

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number  
**WO 03/102923 A2**

(51) International Patent Classification<sup>7</sup>: **G10L 21/02**,  
19/14

(21) International Application Number: PCT/CA03/00828

(22) International Filing Date: 30 May 2003 (30.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
2,388,352 31 May 2002 (31.05.2002) CA

(71) Applicant (for all designated States except US):  
**VOICEAGE CORPORATION** [CA/CA]; 750 chemin  
Lucerne, Suite 250, Ville Mont-Royal, Québec H3R 2H6  
(CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BESSETTE, Bruno**  
[CA/CA]; 1546 Pérodeau, Rock Forest, Québec J1N 1L2  
(CA). **LAFLAMME, Claude** [CA/CA]; 294 chemin  
Dépôt, Bonsecours, Québec J0E 1H0 (CA). **JELINEK,**  
**Milan** [CA/CA]; 245 Merrill Park, C.P. 556, North Hatley,  
Québec J0B 2C0 (CA). **LEFEBVRE, Roch** [CA/CA];  
259 Avenue de la Bourgade, Canton de Magog, Québec  
J1X 5R9 (CA).

(74) Agents: **BROUILLETTE, Robert** et al.; Brouillette  
Kosie Prince, 1100 René-Lévesque Blvd. West, 25th  
Floor, Montréal, Québec H3B 5C9 (CA).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD,  
SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, YU, ZA, ZM, ZW.

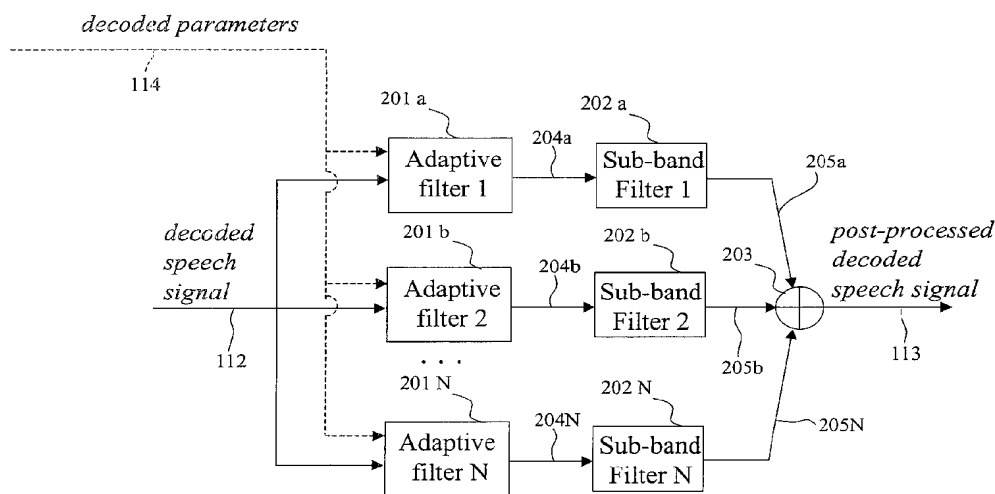
(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,  
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,  
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: A METHOD AND DEVICE FOR FREQUENCY-SELECTIVE PITCH ENHANCEMENT OF SYNTHESIZED  
SPEECH



(57) Abstract: In a method and device for post-processing a decoded sound signal in view of enhancing a perceived quality of this decoded sound signal, the decoded sound signal is divided into a plurality of frequency sub-band signals, and post-processing is applied to at least one of the frequency sub-band signal. After post-processing of this at least one frequency sub-band signal, the frequency sub-band signals may be added to produce an output post-processed decoded sound signal. In this manner, the post-processing can be localized to a desired sub-band or sub-bands with leaving other sub-bands virtually unaltered.

A METHOD AND DEVICE FOR FREQUENCY-SELECTIVE  
PITCH ENHANCEMENT OF SYNTHESIZED SPEECH

5

BACKGROUND OF THE INVENTION

## 10 1. Field of the invention:

The present invention relates to a method and device for post-processing a decoded sound signal in view of enhancing a perceived quality of this decoded sound signal.

15

These post-processing method and device can be applied, in particular but not exclusively, to digital encoding of sound (including speech) signals. For example, these post-processing method and device can also be applied to the more general case of signal enhancement where the noise source can be from any medium or system, not necessarily related to encoding or quantization noise.

20

## 2. Brief description of the current technology:

25

*2.1 Speech encoders*

30

Speech encoders are widely used in digital communication systems to efficiently transmit and/or store speech signals. In digital systems, the analog input speech signal is first sampled at an appropriate sampling rate, and the successive speech samples are further processed in the digital domain. In particular, a speech encoder receives the speech samples as an input, and

generates a compressed output bit stream to be transmitted through a channel or stored on an appropriate storage medium. At the receiver, a speech decoder receives the bit stream as an input, and produces an output reconstructed speech signal.

5

To be useful, a speech encoder must produce a compressed bit stream with a bit rate lower than the bit rate of the digital, sampled input speech signal. State-of-the-art speech encoders typically achieve a compression ratio of at least 16 to 1 and still enable the decoding of high quality speech. Many of  
10 these state-of-the-art speech encoders are based on the CELP (Code-Excited Linear Predictive) model, with different variants depending on the algorithm.

In CELP encoding, the digital speech signal is processed in successive blocks of speech samples called *frames*. For each frame, the encoder extracts  
15 from the digital speech samples a number of parameters that are digitally encoded, and then transmitted and/or stored. The decoder is designed to process the received parameters to reconstruct, or synthesize the given frame of speech signal. Typically, the following parameters are extracted from the digital speech samples by a CELP encoder:

- 20 - Linear Prediction Coefficients (LP coefficients), transmitted in a transformed domain such as the Line Spectral Frequencies (LSF) or Immitance Spectral Frequencies (ISF);
- Pitch parameters, including a pitch delay (or lag) and a pitch gain; and
- Innovative excitation parameters (fixed codebook index and gain).

25 The pitch parameters and the innovative excitation parameters together describe what is called the *excitation* signal. This excitation signal is supplied as an input to a Linear Prediction (LP) filter described by the LP coefficients. The LP filter can be viewed as a model of the vocal tract, whereas the excitation signal can be viewed as the output of the glottis. The LP or LSF  
30 coefficients are typically calculated and transmitted every frame, whereas the pitch and innovative excitation parameters are calculated and transmitted

several times per frame. More specifically, each frame is divided into several signal blocks called *subframes*, and the pitch parameters and the innovative excitation parameters are calculated and transmitted every subframe. A frame typically has a duration of 10 to 30 milliseconds, whereas a subframe typically  
5 has a duration of 5 milliseconds.

Several speech encoding standards are based on the Algebraic CELP (ACELP) model, and more precisely on the ACELP algorithm. One of the main features of ACELP is the use of algebraic codebooks to encode the innovative  
10 excitation at each subframe. An algebraic codebook divides a subframe in a set of *tracks* of interleaved pulse positions. Only a few non-zero-amplitude pulses per track are allowed, and each non-zero-amplitude pulse is restricted to the positions of the corresponding track. The encoder uses fast search algorithms to find the optimal pulse positions and amplitudes for the pulses of  
15 each subframe. A description of the ACELP algorithm can be found in the article of R. SALAMI et al., "*Design and description of CS-ACELP: a toll quality 8 kb/s speech coder*", *IEEE Trans. on Speech and Audio Proc.*, Vol. 6, No. 2, pp. 116-130, March 1998, herein incorporated by reference, and which describes the ITU-T G.729 CS-ACELP narrowband speech encoding algorithm  
20 at 8 kbits/second. It should be noted that there are several variations of the ACELP innovation codebook search, depending on the standard of concern. The present invention is not dependent on these variations, since it only applies to post-processing of the decoded (synthesized) speech signal.

25 A recent standard based on the ACELP algorithm is the ETSI/3GPP AMR-WB speech encoding algorithm, which was also adopted by the ITU-T (Telecommunication Standardization Sector of ITU (International Telecommunication Union)) as recommendation G.722.2 [ITU-T Recommendation G.722.2 "*Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)*", Geneva, 2002], [3GPP TS  
30 26.190, "*AMR Wideband Speech Codec: Transcoding Functions*," 3GPP

*Technical Specification*]. The AMR-WB is a multi-rate algorithm designed to operate at nine different bit rates between 6.6 and 23.85 kbits/second. Those of ordinary skill in the art know that the quality of the decoded speech generally increases with the bit rate. The AMR-WB has been designed to allow cellular communication systems to reduce the bit rate of the speech encoder in the case of bad channel conditions; the bits are converted to channel encoding bits to increase the protection of the transmitted bits. In this manner, the overall quality of the transmitted bits can be kept higher than in the case where the speech encoder operates at a single fixed bit rate.

Figure 7 is a schematic block diagram showing the principle of the AMR-WB decoder. More specifically, Figure 7 is a high-level representation of the decoder, emphasizing the fact that the received bitstream encodes the speech signal only up to 6.4 kHz (12.8 kHz sampling frequency), and the frequencies higher than 6.4 kHz are synthesized at the decoder from the lower-band parameters. This implies that, in the encoder, the original wideband, 16 kHz-sampled speech signal was first down-sampled to the 12.8 kHz sampling frequency, using multi-rate conversion techniques well known to those of ordinary skill in the art. The parameter decoder 701 and the speech decoder 702 of Figure 7 are analogous to the parameter decoder 106 and the source decoder 107 of Figure 1. The received bitstream 709 is first decoded by the parameter decoder 701 to recover parameters 710 supplied to the speech decoder 702 to resynthesize the speech signal. In the specific case of the AMR-WB decoder, these parameters are:

- ISF coefficients for every frame of 20 milliseconds;
- An integer pitch delay  $T_0$ , a fractional pitch value  $T_{0\_frac}$  around  $T_0$ , and a pitch gain for every 5 millisecond subframe; and
- An algebraic codebook shape (pulse positions and signs) and gain for every 5 millisecond subframe.

From the parameters 710, the speech decoder 702 is designed to synthesize a given frame of speech signal for the frequencies equal to and lower than 6.4

kHz, and thereby produce a low-band synthesized speech signal 712 at the 12.8 kHz sampling frequency. To recover the full-band signal corresponding to the 16 kHz sampling frequency, the AMR-WB decoder comprises a high-band resynthesis processor 707 responsive to the decoded parameters 710 from the parameter decoder 701 to resynthesize a high-band signal 711 at the sampling frequency of 16 kHz. The details of the high-band signal resynthesis processor 707 can be found in the following publications which are herein incorporated by reference:

- 10 - *ITU- T Recommendation G.722.2 "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", Geneva, 2002; and*
- 15 - *3GPP TS 26.190, "AMR Wideband Speech Codec: Transcoding Functions," 3GPP Technical Specification.*

The output of the high-band resynthesis processor 707, referred to as the high-band signal 711 of Figure 7, is a signal at the 16 kHz sampling frequency, having an energy concentrated above 6.4 kHz. The processor 708 sums the high-band signal 711 to a 16-kHz up-sampled low-band speech signal 713 to form the complete decoded speech signal 714 of the AMR-WB decoder at the 16 kHz sampling frequency.

## 2.2 Need for post-processing

25 Whenever a speech encoder is used in a communication system, the synthesized or decoded speech signal is never identical to the original speech signal even in the absence of transmission errors. The higher the compression ratio, the higher the distortion introduced by the encoder. This distortion can be made subjectively small using different approaches. A first approach is to condition the signal at the encoder to better describe, or encode, subjectively

relevant information in the speech signal. The use of a formant weighting filter, often represented as  $W(z)$ , is a widely used example of this first approach [B. Kleijn and K. Paliwal editors, «Speech Coding and Synthesis, » Elsevier, 1995]. This filter  $W(z)$  is typically made adaptive, and is computed in such a way that it reduces the signal energy near the spectral formants, thereby increasing the relative energy of lower energy bands. The encoder can then better quantize lower energy bands, which would otherwise be masked by encoding noise, increasing the perceived distortion. Another example of signal conditioning at the encoder is the so-called *pitch sharpening* filter which enhances the harmonic structure of the excitation signal at the encoder. Pitch sharpening aims at ensuring that the inter-harmonic noise level is kept low enough in the perceptual sense.

A second approach to minimize the perceived distortion introduced by a speech encoder is to apply a so-called *post-processing* algorithm. Post-processing is applied at the decoder, as shown in Figure 1. In Figure 1, the speech encoder 101 and the speech decoder 105 are broken down in two modules. In the case of the speech encoder 101, a source encoder 102 produces a series of speech encoding parameters 109 to be transmitted or stored. These parameters 109 are then binary encoded by the parameter encoder 103 using a specific encoding method, depending on the speech encoding algorithm and on the parameters to encode. The encoded speech signal (binary encoded parameters) 110 is then transmitted to the decoder through a communication channel 104. At the decoder, the received bit stream 111 is first analysed by a parameter decoder 106 to decode the received, encoded sound signal encoding parameters, which are then used by the source decoder 107 to generate the synthesized speech signal 112. The aim of post-processing (see post-processor 108 of Figure 1) is to enhance the perceptually relevant information in the synthesized speech signal, or equivalently to reduce or remove the perceptually annoying information. Two commonly used forms of post-processing are formant post-processing and

pitch post-processing. In the first case, the formant structure of the synthesized speech signal is amplified by the use of an adaptive filter with a frequency response correlated to the speech formants. The spectral peaks of the synthesized speech signal are then accentuated at the expense of spectral valleys whose relative energy becomes smaller. In the case of pitch post-processing, an adaptive filter is also applied to the synthesized speech signal. However in this case, the filter's frequency response is correlated to the fine spectral structure, namely the harmonics. A pitch post-filter then accentuates the harmonics at the expense of inter-harmonic energy which becomes relatively smaller. Note that the frequency response of a pitch post-filter typically covers the whole frequency range. The impact is that a harmonic structure is imposed on the post-processed speech even in frequency bands that did not exhibit a harmonic structure in the decoded speech. This is not a perceptually optimal approach for wideband speech (speech sampled at 16 kHz), which rarely exhibits a periodic structure on the whole frequency range.

#### SUMMARY OF THE INVENTION

20

The present invention relates to a method for post-processing a decoded sound signal in view of enhancing a perceived quality of this decoded sound signal, comprising dividing the decoded sound signal into a plurality of frequency sub-band signals, and applying post-processing to at least one of the frequency sub-band signals, but not all the frequency sub-band signals.

25

The present invention is also concerned with a device for post-processing a decoded sound signal in view of enhancing a perceived quality of this decoded sound signal, comprising means for dividing the decoded sound signal into a plurality of frequency sub-band signals, and means for post-processing at least one of the frequency sub-band signals, but not all the

30



frequency sub-band signals.

According to an illustrative embodiment, after post-processing of the above mentioned at least one frequency sub-band signal, the frequency sub-band signals are summed to produce an output post-processed decoded sound signal.

Accordingly, the post-processing method and device make it possible to localize the post-processing in the desired sub-band(s) and to leave other sub-bands virtually unaltered.

The present invention further relates to a sound signal decoder comprising an input for receiving an encoded sound signal, a parameter decoder supplied with the encoded sound signal for decoding sound signal encoding parameters, a sound signal decoder supplied with the decoded sound signal encoding parameters for producing a decoded sound signal, and a post processing device as described above for post-processing the decoded sound signal in view of enhancing a perceived quality of this decoded sound signal.

20

The foregoing and other objects, advantages and features of the present invention will become more apparent upon reading of the following, non restrictive description of illustrative embodiments thereof, given by way of example only with reference to the accompanying drawings.

25

#### BRIEF DESCRIPTION OF THE DRAWINGS

30 In the appended drawings:

Figure 1 is a schematic block diagram of the high-level structure of an example of speech encoder/decoder system using post-processing at the decoder;

5        Figure 2 is a schematic block diagram showing the general principle of an illustrative embodiment of the present invention using a bank of adaptive filters and sub-band filters, in which the input of the adaptive filters is the decoded (synthesized) speech signal (solid line) and the decoded parameters (dotted line);

10

Figure 3 is a schematic block diagram of a two-band pitch enhancer, which constitutes a special case of the illustrative embodiment of Figure 2;

15        Figure 4 is a schematic block diagram of an illustrative embodiment of the present invention, as applied to the special case of the AMR-WB wideband speech decoder;

20        Figure 5 is a schematic block diagram of an alternative implementation of the illustrative embodiment of Figure 4;

20

Figure 6a is a graph illustrating an example of spectrum of a pre-processed signal;

25        Figure 6b is a graph illustrating an example of spectrum of the post-processed signal obtained when using the method described in Figure 3;

Figure 7 is a schematic block diagram showing the principle of operation of the 3GPP AMR-WB decoder;

30

Figures 8a and 8b are graphs showing an example of the frequency response of a pitch enhancer filter as described by Equation (1), with the

special case of a pitch period  $T=10$  samples;

Figure 9a is a graph showing an example of frequency response for the low-pass filter 404 of Figure 4;

5

Figure 9b is a graph showing an example of frequency response for the band-pass filter 407 of Figure 4;

Figure 9c is a graph showing an example of combined frequency response for the low-pass filter 404 and band-pass filters 407 of Figure 4; and

10

Figure 10 is a graph showing an example of the frequency response of an inter-harmonic filter as described by Equation (2), and used in the inter-harmonic filter 503 of Figure 5, for the specific case of  $T=10$  samples.

15

#### DETAILED DESCRIPTION OF THE ILLUSTRATIVE EMBODIMENTS

Figure 2 is a schematic block diagram illustrating the general principle of an illustrative embodiment of the present invention.

20

In Figure 1, the input signal (signal on which post-processing is applied) is the decoded (synthesized) speech signal 112 produced by the speech decoder 105 (Figure 1) at the receiver of a communications system (output of the source decoder 107 of Figure 1). The aim is to produce a post-processed decoded speech signal at the output 113 of the post-processor 108 of Figure 1 (which is also the output of processor 203 of Figure 2) with enhanced perceived quality. This is achieved by first applying at least one, and possibly more than one, adaptive filtering operation to the input signal 112 (see adaptive filters 201a, 201b, ... , 201N). These adaptive filters will be described

25

30

in the following description. It should be pointed out here that some of the adaptive filters 201a to 201N can be trivial functions whenever required, for example with the output equal to the input. The output 204a, 204b, ... , 204N of each adaptive filter 201a, 201b, ..., 201N is then band-pass filtered through a sub-band filter 202a, 202b, ... , 202N, respectively, and the post-processed decoded speech signal 113 is obtained by adding through a processor 203 the respective resulting outputs 205a, 205b, ... , 205N of sub-band filters 202a, 202b, ... , 202N.

10 In one illustrative embodiment, a two-band decomposition is used and adaptive filtering is applied only to the lower band. This results in a total post-processing that is mostly targeted at frequencies near the first harmonics of the synthesized speech signal.

15 Figure 3 is a schematic block diagram of a two-band pitch enhancer, which constitutes a special case of the illustrative embodiment of Figure 2. More specifically, Figure 3 shows the basic functions of a two-band post-processor (see post-processor 108 of Figure 1). According to this illustrative embodiment, only pitch enhancement is considered as post-processing although other types of post-processing could be contemplated. In Figure 3, the decoded speech signal (assumed to be the output 112 of the source decoder 107 of Figure 1) is supplied through a pair of sub-branches 308 and 309.

25 In the higher branch 308, the decoded speech signal 112 is filtered by a high-pass filter 301 to produce the higher band signal 310 ( $s_H$ ). In this specific example, no adaptive filter is used in the higher branch. In the lower branch 309, the decoded speech signal 112 is first processed through an adaptive filter 307 comprising an optional low-pass filter 302, a pitch tracking module 303, and a pitch enhancer 304, and then filtered through a low-pass filter 305 to obtain the lower band, post processed signal 311 ( $s_{LEF}$ ). The post-

processed decoded speech signal 113 is obtained by adding through an adder 306 the lower 311 and higher 312 band post-processed signals from the output of the low-pass filter 305 and high-pass filter 301, respectively. It should be pointed out that the low-pass 305 and high-pass 301 filters could be of  
 5 many different types, for example Infinite Impulse Response (UR) or Finite Impulse Response (FIR). In this illustrative embodiment, linear phase FIR filters are used.

Therefore, the adaptive filter 307 of Figure 3 is composed of two, and  
 10 possibly three processors, the optional low-pass filter 302 similar to low-pass filter 305, the pitch tracking module 303 and the pitch enhancer 304.

The low-pass filter 302 can be omitted, but it is included to allow viewing of the post-processing of Figure 3 as a two-band decomposition  
 15 followed by specific filtering in each sub-band. After optional low-pass filtering (filter 302) of the decoded speech signal 112 in the lower band, the resulting signal  $s_L$  is processed through the pitch enhancer 304. The object of the pitch enhancer 304 is to reduce the inter-harmonic noise in the decoded speech signal. In the present illustrative embodiment, the pitch enhancer 304 is  
 20 achieved by a time-varying linear filter described by the following equation :

$$y(n) = \left(1 - \frac{\alpha}{2}\right)x[n] + \frac{\alpha}{4}\{x[n-T] + x[n+T]\} \quad (1)$$

where  $\alpha$  is a coefficient that controls the inter-harmonic attenuation,  $T$  is the  
 25 pitch period of the input signal  $x[n]$ , and  $y[n]$  is the output signal of the pitch enhancer. A more general equation could also be used where the filter taps at  $n-T$  and  $n+T$  could be at different delays (for example  $n-T1$  and  $n+T2$ ). Parameters  $T$  and  $\alpha$  vary with time and are given by the pitch tracking module 303. With a value of  $\alpha = 1$ , the gain of the filter described by Equation (1) is  
 30 exactly 0 at frequencies  $1/(2T), 3/(2T), 5/(2T)$ , etc, i.e. at the mid-point between

the harmonic frequencies  $1/T$ ,  $3/T$ ,  $5/T$ , etc. When  $\alpha$  approaches 0, the attenuation between the harmonics produced by the filter of Equation (1) reduces. With a value of  $\alpha = 0$ , the filter output is equal to its input. Figure 8 shows the frequency response (in dB) of the filter described by Equation (1) for the values  $\alpha = 0.8$  and 1, when the pitch delay is (arbitrarily) set at a value  $T = 10$  samples. The value of  $\alpha$  can be computed using several approaches. For example, the normalized pitch correlation, which is well-known by those of ordinary skill in the art, can be used to control the coefficient  $\alpha$ : the higher the normalized pitch correlation (the closer to 1 it is), the higher the value of  $\alpha$ . A periodic signal  $x[n]$  with a period of  $T = 10$  samples would have harmonics at the maxima of the frequency responses of Figure 8, i.e. at normalized frequencies 0.2, 0.4, etc. It is easy to understand from Figure 8 that the pitch enhancer of Equation (1) would attenuate the signal energy only between its harmonics, and that the harmonic components would not be altered by the filter. Figure 8 also shows that varying parameter  $\alpha$  enables control of the amount of inter-harmonic attenuation provided by the filter of Equation (1). Note that the frequency response of the filter of Equation (1), shown in Figure 8, extends to all frequencies of the spectrum.

Since the pitch period of a speech signal varies in time, the pitch value  $T$  of the pitch enhancer 304 has to vary accordingly. The pitch tracking module 303 is responsible for providing the proper pitch value  $T$  to the pitch enhancer 304, for every frame of the decoded speech signal that has to be processed. For that purpose, the pitch tracking module 303 receives as input not only the decoded speech samples but also the decoded parameters 114 from the parameter decoder 106 of Figure 1.

Since a typical speech encoder extracts, for every speech subframe, a pitch delay which we call  $T_0$  and possibly a fractional value  $T_{0\_frac}$  used to interpolate the adaptive codebook contribution to fractional sample resolution, the pitch tracking module 303 can then use this decoded pitch delay to focus

the pitch tracking at the decoder. One possibility is to use  $T_0$  and  $T_{0\_frac}$  directly in the pitch enhancer 304, exploiting the fact that the encoder has already performed pitch tracking. Another possibility, used in this illustrative embodiment, is to recalculate the pitch tracking at the decoder focussing on  
 5 values around, and multiples or submultiples of, the decoded pitch value  $T_0$ . The pitch tracking module 303 then provides a pitch delay  $T$  to the pitch enhancer 304, which uses this value of  $T$  in Equation (1) for the present frame of decoded speech signal. The output is signal  $s_{LE}$ .

10 Pitch enhanced signal  $s_{LE}$  is then low-pass filtered through filter 305 to isolate the low frequencies of the pitch enhanced signal  $s_{LE}$ , and to remove the high-frequency components that arise when the pitch enhancer filter of Equation (1) is varied in time, according to the pitch delay  $T$ , at the decoded speech frame boundaries. This produces the lower band post-processed  
 15 signal  $s_{LEF}$ , which can now be added to the higher band signal  $s_H$  in the adder 306. The result is the post-processed decoded speech signal 113, with reduced inter-harmonic noise in the lower band. The frequency band where pitch enhancement will be applied depends on the cut-off frequency of the low-pass filter 305 (and optionally in low-pass filter 302).

20  
 Figures 6a and 6b show an example signal spectrum illustrating the effect of the post-processing described in Figure 3. Figure 6a is the spectrum of the input signal 112 of the post-processor 108 of Figure 1 (decoded speech signal 112 in Figure 3). In this illustrative example, the input signal is  
 25 composed of 20 harmonics, with fundamental frequency  $f_0 = 373$  Hz chosen arbitrarily, with «noisy» components added at frequencies  $f_0/2$ ,  $3f_0/2$  and  $5f_0/2$ . These three noisy components can be seen between the low-frequency harmonics in Figure 6a. The sampling frequency is assumed to be 16 kHz in this example. The two-band pitch enhancer shown in Figure 3 and described  
 30 above is then applied to the signal of Figure 6a. With a sampling frequency of 16 kHz and a periodic signal of fundamental frequency equal to 373 Hz as in

Figure 6a, the pitch tracking module 303 should find a period of  $T = 16000/373 \approx 43$  samples. This is the value that was used for the pitch enhancer filter of Equation (1), applied to the pitch enhancer 304 of Figure 3. A value of  $\alpha = 0.5$  was also used. The low-pass 305 and high-pass 301 filters are symmetric, linear phase FIR filters with 31 taps. The cut-off frequency for this example is chosen as 2000 Hz. These specific values are given only as an illustrative example.

The post-processed decoded speech signal 113 at the output of the adder 306 has a spectrum shown in Figure 6b. It can be seen that the three inter-harmonic sinusoids in Figure 6a have been completely removed, while the harmonics of the signal have been practically unaltered. Also it is noted that the effect of the pitch enhancer diminishes as the frequency approaches the low-pass filter cut-off frequency (2000 Hz in this example). Hence, only the lower band is affected by the post-processing. This is a key feature of this illustrative embodiment of the present invention. By varying the cut-off frequencies of the optional low-pass filter 302, low-pass filter 305 and high-pass filter 301, it is possible to control up to which frequency pitch enhancement is applied.

#### *Application to the AMR-WB speech decoder*

The present invention can be applied to any speech signal synthesized by a speech decoder, or even to any speech signal corrupted by inter-harmonic noise that needs to be reduced. This section will show a specific, exemplary implementation of the present invention to an AMR-WB decoded speech signal. The post-processing is applied to the low-band synthesized speech signal 712 of Figure 7, i.e. to the output of the speech decoder 702, which produces a synthesized speech at a sampling frequency of 12.8 kHz.

Figure 4 shows the block diagram of a pitch post-processor when the



input signal is the AMR-WB low-band synthesized speech signal at the sampling frequency of 12.8 kHz. More precisely, the post-processor presented in Figure 4 replaces the up-sampling unit 703, which comprises processors 704, 705 and 706. The pitch post-processor of Figure 4 could also be applied to the 16 kHz up-sampled synthesized speech signal, but applying it prior to up-sampling reduces the number of filtering operations at the decoder, and thus reduces complexity.

The input signal (*AMR-WB low-band synthesized speech (12.8 kHz)*) of Figure 4 is designated as signal  $s$ . In this specific example, signal  $s$  is the AMR-WB low-band synthesized speech signal at the sampling frequency of 12.8 kHz (output of processor 702). The pitch post-processor of Figure 4 comprises a pitch tracking module 401 to determine, for every 5 millisecond subframe, the pitch delay  $T$  using the received, decoded parameters 114 (Figure 1) and the synthesized speech signal  $s$ . The decoded parameters used by the pitch tracking module are  $T_0$ , the integer pitch value for the subframe, and  $T_{0\_frac}$ , the fractional pitch value for subsample resolution. The pitch delay  $T$  calculated in the pitch tracking module 401 will be used in the next steps for pitch enhancement. It would be possible to use directly the received, decoded pitch parameters  $T_0$  and  $T_{0\_frac}$  to form the delay  $T$  used by the pitch enhancer in the pitch filter 402. However, the pitch tracking module 401 is capable of correcting pitch multiples or submultiples, which could have a harmful effect on the pitch enhancement.

An illustrative embodiment of pitch tracking algorithm for the module 401 is the following (the specific thresholds and pitch tracked values are given only by way of example):

- First, the decoded pitch information (pitch delay  $T_0$ ) is compared to a stored value of the decoded pitch delay  $T_{prev}$  of the previous frame.  $T_{prev}$  may have been modified by some of the following steps

according to the pitch tracking algorithm. For example, if  $T_0 < 1.16 \cdot T_{prev}$  then go to case 1 below, else if  $T_0 > 1.16 \cdot T_{prev}$ , then set  $T_{temp} = T_0$  and go to case 2 below.

5                   Case 1: First, calculate the cross-correlation  $C2$  (cross-product) between the last synthesized subframe and the synthesis signal starting at  $T_0/2$  samples before the beginning of the last subframe (look at correlation at half the decoded pitch value).

10                   Then, calculate the cross-correlation  $C3$  (cross-product) between the last synthesized subframe and the synthesis signal starting at  $T_0/3$  samples before the beginning of the last subframe (look at correlation at one-third the decoded pitch value).

15                   Then, select the maximum value between  $C2$  and  $C3$  and calculate the normalized correlation  $Cn$  (normalized version of  $C2$  or  $C3$ ) at the corresponding sub-multiple of  $T_0$  (at  $T_0/2$  if  $C2 > C3$  and at  $T_0/3$  if  $C3 > C2$ ). Call  $T_{new}$  the pitch sub-multiple corresponding to the highest normalized correlation.

20                   If  $Cn > 0.95$  (strong normalized correlation) the new pitch period is  $T_{new}$  (instead of  $T_0$ ). Output the value  $T = T_{new}$  from the pitch tracking module 401. Save  $T_{prev} = T$  for next subframe pitch tracking and exit the pitch tracking module 401.

25                   If  $0.7 < Cn < 0.95$ , then save  $T_{temp} = T_0/2$  or  $T_0/3$  (according to  $C2$  or  $C3$  above) for comparisons in case 2

below. Otherwise, if  $C_n < 0.7$  save  $T\_temp = T_0$ .

Case 2: Calculate all possible values of the ratio  $T_n = [T\_temp/n]$  where  $[x]$  means the integer part of  $x$  and  $n = 1, 2, 3$ , etc. is an integer.

5

10

Calculate all cross correlations  $C_n$  at the pitch delay submultiples  $T_n$ . Retain  $C_{n\_max}$  as the maximum cross correlation among all  $C_n$ . If  $n > 1$  and  $C_n > 0.8$ , output  $T_n$  as the pitch period output  $T$  of the pitch tracking unit 401. Otherwise, output  $T1 = T\_temp$ . Here, the value of  $T\_temp$  will depend on the calculations in Case 1 above.

It should be noted that the above example of pitch tracking module 401 is given for the purpose of illustration only. Any other pitch tracking method or device could be implemented in module 401 (or 303 and 502) to ensure a better pitch tracking at the decoder.

Therefore, the output of the pitch tracking module is the period  $T$  to be used in the pitch filter 402 which, in this preferred embodiment, is described by the filter of Equation (1). Again, a value of  $\alpha = 0$  implies no filtering (output of the pitch filter 402 is equal to its input), and a value of  $\alpha = 1$  corresponds to the highest amount of pitch enhancement.

Once the enhanced signal  $S_E$  (Figure 4) is determined, it is combined with the input signal  $s$  such that, as in Figure 3, only the lower band is subjected to pitch enhancement. In Figure 4, a modified approach is used compared to Figure 3. Since the pitch post-processor of Figure 4 replaces the up-sampling unit 703 in Figure 7, the sub-band filters 301 and 305 of Figure 3 are combined with the interpolation filter 705 of Figure 7 to minimize the number of filtering operations, and the filtering delay. More specifically, filters

30

404 and 407 of Figure 4 act both as band-pass filters (to separate the frequency bands) and as interpolation filters (for up-sampling from 12.8 to 16 kHz). These filters 404 and 407 could be further designed such that the band-pass filter 407 has relaxed constraints in its low-frequency stop band (i.e. it does not have to completely attenuate the signal at low frequencies). This could be achieved by using design constraints similar to those shown in Figure 9. Figure 9a is an example of frequency response for the low-pass filter 404. It should be noted that the DC (Direct Current) gain of this filter is 5 (instead of 1) since this filter also acts as interpolation filter, with a 5/4 interpolation ratio which implies that the filter gain must be 5 at 0 Hz. Then, Figure 9b shows the frequency response of the band-pass filter 407 making this filter 407 complementary, in the low band, to the low-pass filter 404. In this example, the filter 407 is a band-pass filter, not a high-pass filter such as filter 301, since it must act both as high-pass filter (such as filter 301) and low-pass filter (such as interpolation filter 705). Referring again to Figure 9, we see that the low-pass and band-pass filters 404 and 407 are complementary when considered in parallel, as in Figure 4. Their combined frequency response (when used in parallel) is shown in Figure 9c.

For completeness, the tables of filter coefficients used in this illustrative embodiment of the filters 404 and 407 are given below. Of course, these tables of filter coefficients are given by way of example only. It should be understood that these filters can be replaced without modifying the scope, spirit and nature of the present invention.

Table 1. Low-pass coefficients of filter 404

hlp[0]	0.04375000000000	hlp[30]	0.01998000000000
hlp[1]	0.04371500000000	hlp[31]	0.01882400000000
hlp[2]	0.04361200000000	hlp[32]	0.01768200000000
hlp[3]	0.04344000000000	hlp[33]	0.01655700000000

hlp[4]	0.04320000000000	hlp[34]	0.01545100000000
hlp[ 5]	0.04289300000000	hlp[35]	0.01436900000000
hlp[ 6]	0.04252100000000	hlp[36]	0.01331200000000
hlp[7]	0.04208300000000	hlp[37]	0.01228400000000
hlp[8]	0.04158200000000	hlp[38]	0.01128600000000
hlp[9]	0.04102000000000	hlp[39]	0.01032300000000
hlp[10]	0.04039900000000	hlp[ 40]	0.00939500000000
hlp[11]	0.03972100000000	hlp[41]	0.00850500000000
hlp[12]	0.03898800000000	hlp[ 42]	0.00765500000000
hlp[13]	0.03820200000000	hlp[ 43]	0.00684600000000
hlp[14]	0.03736700000000	hlp[ 44]	0.00608100000000
hlp[15]	0.03648600000000	hlp[ 45]	0.00535900000000
hlp[ 16]	0.03556100000000	hlp[ 46]	0.00468200000000
hlp[17]	0.03459600000000	hlp[ 47]	0.00405100000000
hlp[18]	0.03359400000000	hlp[ 48]	0.00346700000000
hlp[19]	0.03255800000000	hlp[ 49]	0.00292900000000
hlp[20]	0.03149200000000	hlp[50]	0.00243900000000
hlp[21]	0.03039900000000	hlp[51]	0.00199500000000
hlp[22]	0.02928400000000	hlp[ 52]	0.00159900000000
hlp[23]	0.02814900000000	hlp[53]	0.00124800000000
hlp[24]	0.02699900000000	hlp[54]	0.00094400000000
hlp[25]	0.02583700000000	hlp[55]	0.00068400000000
hlp[26]	0.02466700000000	hlp[56]	0.00046800000000
hlp[27]	0.02349300000000	hlp[ 57]	0.00029500000000
hlp[28]	0.02231800000000	hlp[58]	0.00016300000000
hlp[29]	0.02114600000000	hlp[59]	0.00007100000000
		hlp[ 60]	0.00001800000000

Table 2. Band-pass coefficients of filter 407

hbp[0]	0.956250000000000	hbp[30]	-0.019980000000000
hbp[1]	0.891154000000000	hbp[31]	-0.004124000000000
hbp[2]	0.711209000000000	hbp[32]	0.004143000000000
hbp[3]	0.458106000000000	hbp[33]	0.003433000000000
hbp[4]	0.188199000000000	hbp[34]	-0.004161000000000
hbp[ 5]	-0.042893000000000	hbp[35]	-0.014369000000000
hbp[ 6]	-0.194743000000000	hbp[36]	-0.022673000000000
hbp[7]	-0.251369000000000	hbp[37]	-0.026018000000000
hbp[8]	-0.222872000000000	hbp[38]	-0.023700000000000
hbp[9]	-0.139480000000000	hbp[39]	-0.017232000000000
hbp[10]	-0.040399000000000	hbp[ 40]	-0.009395000000000
hbp[ 11]	0.038681000000000	hbp[41]	-0.002970000000000
hbp[12]	0.075484000000000	hbp[ 42]	0.000305000000000
hbp[13]	0.065665000000000	hbp[ 43]	0.000190000000000
hbp[14]	0.021138000000000	hbp[ 44]	-0.002260000000000
hbp[15]	-0.036486000000000	hbp[ 45]	-0.005359000000000
hbp[16]	-0.084653000000000	hbp[ 46]	-0.007568000000000
hbp[17]	-0.107634000000000	hbp[ 47]	-0.008058000000000
hbp[18]	-0.100876000000000	hbp[ 48]	-0.006870000000000
hbp[19]	-0.070919000000000	hbp[49]	-0.004695000000000
hbp[20]	-0.031492000000000	hbp[ 50]	-0.002439000000000
hbp[21]	0.002342000000000	hbp[51]	-0.000806000000000
hbp[22]	0.019700000000000	hbp[52]	-0.000063000000000
hbp[23]	0.017153000000000	hbp[53]	-0.000053000000000
hbp[24]	-0.001107000000000	hbp[54]	-0.000387000000000
hbp[25]	-0.025837000000000	hbp[55]	-0.000684000000000
hbp[26]	-0.046789000000000	hbp[ 56]	-0.000744000000000
hbp[27]	-0.056549000000000	hbp[57]	-0.000576000000000
hbp[28]	-0.052818000000000	hbp[58]	-0.000319000000000
hbp[29]	-0.038519000000000	hbp[59]	-0.000113000000000

		hbp[60]	-0.00001800000000
--	--	---------	-------------------

The output of the pitch filter 402 of Figure 4 is called  $S_E$ . To be recombined with the signal of the upper branch, it is first up-sampled by processor 403, low-pass filter 404 and processor 405, and added through an adder 409 to the up-sampled upper branch signal 410. The up-sampling operation in the upper branch is performed by processor 406, band-pass filter 407 and processor 408.

#### *Alternate implementation of the proposed pitch enhancer*

10

Figure 5 shows an alternative implementation of a two-band pitch enhancer according to an illustrative embodiment of the present invention. It should be noted that the upper branch of Figure 5 does not process the input signal at all. This means that, in this particular case, the filters in the upper branch of Figure 2 (adaptive filters 201a and 201b) have trivial input-output characteristics (output is equal to input). In the lower branch, the input signal (signal to be enhanced) is processed first through an optional low-pass filter 501, then through a linear filter called inter-harmonic filter 503, defined by the following equation:

20

$$y[n] = \frac{1}{2}x[n] - \frac{1}{4}\{x[n-T] + x[n+T]\} \quad (2)$$

It should be noted that the negative sign in front of the second term on the right hand side, compared to Equation (1). It should also be noted that the enhancement factor  $\alpha$  is not included in Equation (2), but rather it is introduced by means of an adaptive gain by the processor 504 of Figure 5. The inter-harmonic filter 503, described by Equation (2), has a frequency response such that it completely removes the harmonics of a periodic signal having a period of  $T$  samples, and such that a sinusoid at a frequency exactly between the

25

harmonics passes through the filter unchanged in amplitude but with a phase reversal of exactly 180 degrees (same as sign inversion). For example, Figure 10 shows the frequency response of the filter described by Equation (2) when the period is (arbitrarily) chosen at  $T = 10$  samples. A periodic signal with period  $T = 10$  samples would present harmonics at normalized frequencies 0.2, 0.4, 0.6, etc., and Figure 10 shows that the filter of Equation (2), with  $T = 10$  samples, would completely remove these harmonics. On the other hand, the frequencies at the exact mid-point between the harmonics would appear at the output of the filter with the same amplitude but with a 180° phase shift.

5 This is the reason why the filter described by Equation (2) and used as filter 503 is called inter-harmonic filter.

The pitch value  $T$  for use in the inter-harmonic filter 503 is obtained adaptively by the pitch tracking module 502. Pitch tracking module 502 operates on the decoded speech signal and the decoded parameters, similarly to the previously disclosed methods as shown in Figures 3 and 4.

15

Then, the output 507 of the inter-harmonic filter 503 is a signal formed essentially of the inter-harmonic portion of the input decoded signal 112, with 180° phase shift at mid-point between the signal harmonics. Then, the output 507 of the inter-harmonic filter 503 is multiplied by a gain  $\alpha$  (processor 504) and subsequently low-pass filtered (filter 505) to obtain the low frequency band modification that is applied to the input decoded speech signal 112 of Figure 5, to obtain the post-processed decoded signal (enhanced signal) 509. The coefficient  $\alpha$  in processor 504 controls the amount of pitch or inter-harmonic enhancement. The closer to 1 is  $\alpha$ , the higher the enhancement is. When  $\alpha$  is equal to 0, no enhancement is obtained, i.e. the output of adder 506 is exactly equal to the input signal (decoded speech in Figure 5). The value of  $\alpha$  can be computed using several approaches. For example, the normalized pitch correlation, which is well known to those of ordinary skill in the art, can be used to control coefficient  $\alpha$ : the higher the normalized pitch correlation (the

20

25

30



closer to 1 it is), the higher the value of  $\alpha$ .

The final post-processed decoded speech signal 509 is obtained by adding through an adder 506 the output of low-pass filter 505 to the input  
5 signal (decoded speech signal 112 of Figure 5). Depending on the cut-off frequency of the low-pass filter 505, the impact of this post-processing will be limited to the low frequencies of the input signal 112, up to a given frequency. The higher frequencies will be effectively unaffected by the post-processing.

10 *One-band alternative using an adaptive high-pass filter*

One last alternative for implementing sub-band post-processing for enhancing the synthesis signal at low frequencies is to use an adaptive high-pass filter, whose cut-off frequency is varied according to the input signal pitch  
15 value. Specifically, and without referring to any drawing, the low frequency enhancement using this illustrative embodiment would be performed, at each input signal frame, according to the following steps:

- 20 1. Determine the input signal pitch value (signal period) using the input signal and possibly the decoded parameters (output of speech decoder 105) if post-processing a decoded speech signal; this is a similar operation as the pitch tracking operation of modules 303, 401 and 502.
- 25 2. Calculate the coefficients of a high-pass filter such that the cut-off frequency is below, but close to, the fundamental frequency of the input signal; alternatively, interpolate between pre-calculated, stored high-pass filters of known cut-off frequencies (the interpolation can be done in the filter taps domain, or in the pole-zero domain, or in some other transformed domain such as the LSF (Line Spectral Frequencies) of  
30 ISF (Immittance Spectral Frequencies) domain).

3. Filter the input signal frame with the calculated high-pass filter, to obtain the post-processed signal for that frame.

5 It should be pointed out that the present illustrative embodiment of the present invention is equivalent to using only one processing branch in Figure 2, and to define the adaptive filter of that branch as a pitch-controlled high-pass filter. The post-processing achieved with this approach will only affect the frequency range below the first harmonic and not the inter-harmonic energy above the first harmonic.

10

Although the present invention has been described in the foregoing description with reference to illustrative embodiments thereof, these embodiments can be modified at will, within the scope of the appended claims without departing from the spirit and nature of the present invention. For  
15 example, although the illustrative embodiments have been described in relation to a decoded speech signal, those of ordinary skill in the art will appreciate that the concepts of the present invention can be applied to other types of decoded signals, in particular but not exclusively to other types of decoded sound signals.

20

## WHAT IS CLAIMED IS:

1. A method for post-processing a decoded sound signal in view of  
5 enhancing a perceived quality of said decoded sound signal, comprising:  
dividing the decoded sound signal into a plurality of frequency sub-  
band signals; and  
applying post-processing to at least one of the frequency sub-band  
signals, but not all the frequency sub-band signals.
- 10 2. A post-processing method as defined in claim 1, further comprising  
summing the frequency sub-band signals, after post-processing of said at least  
one frequency sub-band signal, to produce an output post-processed decoded  
sound signal.
- 15 3. A post-processing method as defined in claim 1, wherein applying  
post-processing to at least one of the frequency sub-band signals comprises  
adaptively filtering said at least one frequency sub-band signal.
- 20 4. A post-processing method as defined in claim 1, wherein dividing the  
decoded sound signal into a plurality of frequency sub-band signals comprises  
sub-band filtering the decoded sound signal to produce the plurality of  
frequency sub-band signals.
- 25 5. A post-processing method as defined in claim 1, wherein, for said at  
least one of the frequency sub-band signals:  
applying post-processing comprises adaptively filtering the decoded  
sound signal; and  
dividing the decoded sound signal comprises sub-band filtering the  
30 adaptively filtered decoded sound signal.

6. A post-processing method as defined in claim 1, wherein:

dividing the decoded sound signal into a plurality of frequency sub-band signals comprises:

- 5       - high-pass filtering the decoded sound signal to produce a frequency high-band signal; and
- low-pass filtering the decoded sound signal to produce a frequency low-band signal; and

applying post-processing to at least one of the frequency sub-band signals comprises:

- 10       - applying post-processing to the decoded sound signal prior to low-pass filtering the decoded sound signal to produce the frequency low-band signal.

7. A post-processing method as defined in claim 6, wherein applying  
15       post-processing to the decoded sound signal comprises pitch enhancing said decoded sound signal to reduce an inter-harmonic noise in the decoded sound signal.

8. A post-processing method as defined in claim 7, further comprising  
20       low-pass filtering the decoded sound signal prior to pitch enhancing said decoded sound signal.

9. A post-processing method as defined in claim 6, further comprising  
25       summing the frequency high-band and low-band signals to produce an output post-processed decoded sound signal.

10. A post-processing method as defined in claim 1, wherein:

dividing the decoded sound signal into a plurality of frequency sub-band signals comprises:

- 30       - band-pass filtering the decoded sound signal to produce a frequency upper-band signal; and

- low-pass filtering the decoded sound signal to produce a frequency lower-band signal; and

applying post-processing to at least one of the frequency sub-band signals comprises:

- 5     - applying post-processing to the frequency lower-band signal.

11. A post-processing method as defined in claim 10, wherein applying post-processing to the frequency lower-band signal comprises pitch enhancing said frequency lower-band signal prior to low-pass filtering the decoded sound  
10     signal.

12. A post-processing method as defined in claim 10, further comprising summing the frequency upper-band and lower-band signals to produce an output post-processed decoded sound signal.

15

13. A post-processing method as defined in claim 1, wherein:  
dividing the decoded sound signal into a plurality of frequency sub-band signals comprises:

- 20     - low-pass filtering the decoded sound signal to produce a frequency low-band signal; and

applying post-processing to at least one of the frequency sub-band signals comprises:

- applying post-processing to the frequency low-band signal.

25     14. A post-processing method as defined in claim 13, wherein applying post-processing to the frequency low-band signal comprises processing the decoded sound signal through an inter-harmonic filter for inter-harmonic attenuation of the decoded sound signal.

30     15. A post-processing method as defined in claim 14, wherein applying post-processing to the frequency low-band signal comprises multiplying the

inter-harmonic filtered decoded sound signal by an adaptive pitch enhancement gain.

16. A post-processing method as defined in claim 14, further  
5 comprising low-pass filtering the decoded sound signal prior to processing the decoded sound signal through the inter-harmonic filter.

17. A post-processing method as defined in claim 13, further  
comprising summing the decoded sound signal and the frequency low-band  
10 signal to produce an output post-processed decoded sound signal.

18. A post-processing method as defined in claim 13, wherein applying  
post-processing to the frequency low-band signal comprises processing the  
decoded sound signal through an inter-harmonic filter having the following  
15 transfer function:

$$y[n] = \frac{1}{2} x[n] - \frac{1}{4} \{x[n-T] + x[n+T]\}$$

for inter-harmonic attenuation of the decoded sound signal, where  $x[n]$  is the  
20 decoded sound signal,  $y[n]$  is the inter-harmonic filtered decoded sound signal  
in a given sub-band, and  $T$  is a pitch delay of the decoded sound signal.

19. A post-processing method as defined in claim 18, further  
comprising summing the unprocessed decoded sound signal and the inter-  
25 harmonic filtered frequency low-band signal to produce an output post-  
processed decoded sound signal.

20. A post-processing method as defined in claim 1, wherein applying  
post-processing to at least one of the frequency sub-band signals comprises  
30 pitch enhancing the decoded sound signal using the following equation:

$$y[n] = \left(1 - \frac{\alpha}{2}\right)x[n] + \frac{\alpha}{4}\{x[n-T] + x[n+T]\}$$

where  $x[n]$  is the decoded sound signal,  $y[n]$  is the pitch enhanced decoded  
5 sound signal in a given sub-band,  $T$  is a pitch delay of the decoded sound  
signal, and  $\alpha$  is a coefficient varying between 0 and 1 to control an amount of  
inter-harmonic attenuation of the decoded sound signal.

21. A post-processing method as defined in claim 20, comprising  
10 receiving the pitch delay  $T$  through a bitstream.

22. A post-processing method as defined in claim 20, comprising  
decoding the pitch delay  $T$  from a received, encoded bitstream.

15 23. A post-processing method as defined in claim 20, comprising  
calculating the pitch delay  $T$  in response to the decoded sound signal for an  
improved pitch tracking.

24. A post-processing method as defined in claim 1, wherein, during  
20 encoding, the sound signal is down-sampled from a higher sampling frequency  
to a lower sampling frequency, and wherein dividing the decoded sound signal  
into a plurality of frequency sub-band signals comprises up-sampling the  
decoded sound signal from the lower sampling frequency to the higher  
sampling frequency.

25

25. A post-processing method as defined in claim 24, wherein dividing  
the decoded sound signal into a plurality of frequency sub-band signals  
comprises sub-band filtering the decoded sound signal, and wherein the up-  
sampling of the decoded sound signal from the lower sampling frequency to  
30 the higher sampling frequency is combined to the sub-band filtering.

26. A post-processing method as defined in claim 24, comprising:

band-pass filtering the decoded sound signal to produce a frequency upper-band signal, said band-pass filtering of the decoded sound signal being combined with up-sampling of the decoded sound signal from the lower  
5 sampling frequency to the higher sampling frequency; and

post-processing the decoded sound signal and low-pass filtering the post-processed decoded sound signal to produce a frequency lower-band signal, said low-pass filtering of the post-processed decoded sound signal being combined with up-sampling of the post-processed decoded sound signal  
10 from the lower sampling frequency to the higher sampling frequency.

27. A post-processing method as defined in claim 26, further comprising adding the frequency upper-band signal with the frequency lower-band signal to form an output post-processed and up-sampled decoded sound  
15 signal.

28. A post-processing method as defined in claim 26, wherein post-processing of the decoded sound signal comprises pitch enhancing the decoded sound signal to reduce an inter-harmonic noise in the decoded sound  
20 signal.

29. A post-processing method as defined in claim 28, wherein pitch enhancing the decoded sound signal comprises processing the decoded sound signal by means of the following equation:  
25

$$y[n] = \left(1 - \frac{\alpha}{2}\right)x[n] + \frac{\alpha}{4}\{x[n-T] + x[n+T]\}$$

where  $x[n]$  is the decoded sound signal,  $y[n]$  is the pitch enhanced decoded sound signal in a given sub-band,  $T$  is a pitch delay of the decoded sound  
30 signal, and  $\alpha$  is a coefficient varying between 0 and 1 to control an amount of inter-harmonic attenuation of the decoded sound signal.



30. A post-processing method as defined in claim 1, wherein:

dividing the decoded sound signal into a plurality of frequency sub-band signals comprises dividing the decoded sound signal into a frequency upper-band signal and a frequency lower-band signal; and

5 applying post-processing to at least one of the frequency sub-band signals comprises post-processing the frequency lower-band signal.

31. A post-processing method as defined in claim 1, wherein applying  
10 post-processing to said at least one of the frequency sub-band signals comprises:

determining a pitch value of the decoded sound signal;

calculating, in relation to the determined pitch value, a high-pass filter with a cut-off frequency below a fundamental frequency of the decoded sound  
15 signal; and

processing the decoded sound signal through the calculated high-pass filter.

32. A device for post-processing a decoded sound signal in view of  
20 enhancing a perceived quality of said decoded sound signal, comprising:

means for dividing the decoded sound signal into a plurality of frequency sub-band signals; and

means for post-processing at least one of the frequency sub-band signals, but not all the frequency sub-band signals.

25

33. A post-processing device as defined in claim 32, further comprising adder means for summing the frequency sub-band signals, after post-processing of said at least one frequency sub-band signal, to produce an output post-processed decoded sound signal.

30

34. A post-processing device as defined in claim 32, wherein the post-

processing means comprises adaptive filter means supplied with the decoded sound signal.

5        35. A post-processing device as defined in claim 32, wherein the dividing means comprises sub-band filter means supplied with the decoded sound signal.

10        36. A post-processing device as defined in claim 32, wherein, for said at least one of the frequency sub-band signals:  
the post-processing means comprises an adaptive filter supplied with the decoded sound signal to produce an adaptively filtered decoded sound signal; and  
the dividing means comprises a sub-band filter supplied with the adaptively filtered decoded sound signal.

15        37. A post-processing device as defined in claim 32, wherein:  
the dividing means comprises:  
- a high-pass filter supplied with the decoded sound signal to produce a frequency high-band signal; and  
20        - a low-pass filter supplied with the decoded sound signal to produce a frequency low-band signal; and  
the post-processing means comprises:  
- a post-processor for post-processing the decoded sound signal prior to low-pass filtering the decoded sound signal through the low-pass filter.

25        38. A post-processing device as defined in claim 37, wherein the post processor comprises a pitch enhancer supplied with the decoded sound signal to produce a pitch enhanced decoded sound signal.

30        39. A post-processing device as defined in claim 38, further comprising a low-pass filter supplied with the decoded sound signal to produce a low-pass

filtered decoded sound signal supplied to the pitch enhancer.

40. A post-processing device as defined in claim 37, further comprising an adder for summing the frequency high-band and low-band signals to  
5 produce an output post-processed decoded sound signal.

41. A post-processing device as defined in claim 32, wherein:  
the dividing means comprises:

- 10 - a band-pass filter supplied with the decoded sound signal to produce a frequency upper-band signal; and
- a low-pass filter supplied with the decoded sound signal to produce a frequency lower-band signal; and

the post-processing means comprises:

- 15 - a post-processor for post-processing the frequency lower-band signal.

42. A post-processing device as defined in claim 41, wherein the post-processor comprises a pitch filter supplied with the decoded sound signal to produce a pitch enhanced decoded sound signal supplied to the low-pass filter.

20

43. A post-processing device as defined in claim 41, further comprising an adder for summing the frequency upper-band and lower-band signals to produce an output post-processed decoded sound signal.

25

44. A post-processing device as defined in claim 32, wherein:  
the dividing means comprises:

- a low-pass filter supplied with the decoded sound signal to produce a frequency low-band signal; and

the post-processing means comprises:

- 30 - a post-processor for post-processing the decoded sound signal to produce a post-processed decoded sound signal supplied to the low-pass filter.

45. A post-processing device as defined in claim 44, wherein the post-processor comprises an inter-harmonic filter supplied with the decoded sound signal to produce an inter-harmonic, attenuated decoded sound signal.

5

46. A post-processing device as defined in claim 45, wherein the post-processor comprises a multiplier for multiplying the inter-harmonic, attenuated decoded sound signal by an adaptive pitch enhancement gain.

10

47. A post-processing device as defined in claim 45, further comprising a low-pass filter supplied with the decoded sound signal to produce a low-pass filtered decoded sound signal supplied to the inter-harmonic filter.

15

48. A post-processing device as defined in claim 44, further comprising an adder for summing the decoded sound signal and the frequency low-band signal to produce an output post-processed decoded sound signal.

20

49. A post-processing device as defined in claim 44, wherein the post-processor comprises an inter-harmonic filter having the following transfer function:

$$y[n] = \frac{1}{2}x[n] - \frac{1}{4}\{x[n-T] + x[n+T]\}$$

for inter-harmonic attenuating the decoded sound signal, where  $x[n]$  is the decoded sound signal,  $y[n]$  is the inter-harmonic filtered decoded sound signal in a given sub-band, and  $T$  is a pitch delay of the decoded sound signal.

25

50. A post-processing device as defined in claim 49, further comprising an adder for summing the unprocessed decoded sound signal and the inter-harmonic filtered frequency low-band signal to produce an output post-

30

processed decoded sound signal.

51. A post-processing device as defined in claim 32, wherein the post-processing means comprises a pitch enhancer of the decoded sound signal  
5 using the following equation:

$$y[n] = \left(1 - \frac{\alpha}{2}\right)x[n] + \frac{\alpha}{4}\{x[n-T] + x[n+T]\}$$

where  $x[n]$  is the decoded sound signal,  $y[n]$  is the pitch enhanced decoded  
10 sound signal in a given sub-band,  $T$  is a pitch delay of the decoded sound signal, and  $\alpha$  is a coefficient varying between 0 and 1 to control an amount of inter-harmonic attenuation of the decoded sound signal.

52. A post-processing device as defined in claim 51, comprising means  
15 for receiving the pitch delay  $T$  through a bitstream.

53. A post-processing device as defined in claim 51, comprising means  
for decoding the pitch delay  $T$  from a received, encoded bitstream.

20 54. A post-processing device as defined in claim 51, comprising means for calculating the pitch delay  $T$  in response to the decoded sound signal for an improved pitch tracking.

25 55. A post-processing device as defined in claim 32, wherein, during encoding, the sound signal is down-sampled from a higher sampling frequency to a lower sampling frequency, and wherein the dividing means comprises means for up-sampling the decoded sound signal from the lower sampling frequency to the higher sampling frequency.

30 56. A post-processing device as defined in claim 55, wherein the dividing means comprises sub-band filter means supplied with the decoded

sound signal, and wherein the up-sampling means is combined with the sub-band filter means.

57. A post-processing device as defined in claim 55, wherein:

5 - the post-processing means comprises:

means for post-processing the decoded sound signal; and

- the dividing means comprises:

10 a band-pass filter supplied with the decoded sound signal to produce a frequency upper-band signal, said band-pass filter being combined with the up-sampling means; and

a low-pass filter supplied with the post-processed decoded sound signal to produce a frequency lower-band signal, said low-pass filter being combined with the up-sampling means.

15 58. A post-processing device as defined in claim 57, further comprising an adder for summing the frequency upper-band signal with the frequency lower-band signal to form an output post-processed and up-sampled decoded sound signal.

20 59. A post-processing device as defined in claim 57, wherein the means for post-processing the decoded sound signal comprises means for pitch enhancing the decoded sound signal to reduce an inter-harmonic noise in the decoded sound signal.

25 60. A post-processing device as defined in claim 59, wherein the pitch enhancing means comprises means for processing the decoded sound signal by means of the following equation:

$$y[n] = \left(1 - \frac{\alpha}{2}\right)x[n] + \frac{\alpha}{4}\{x[n-T] + x[n+T]\}$$

30

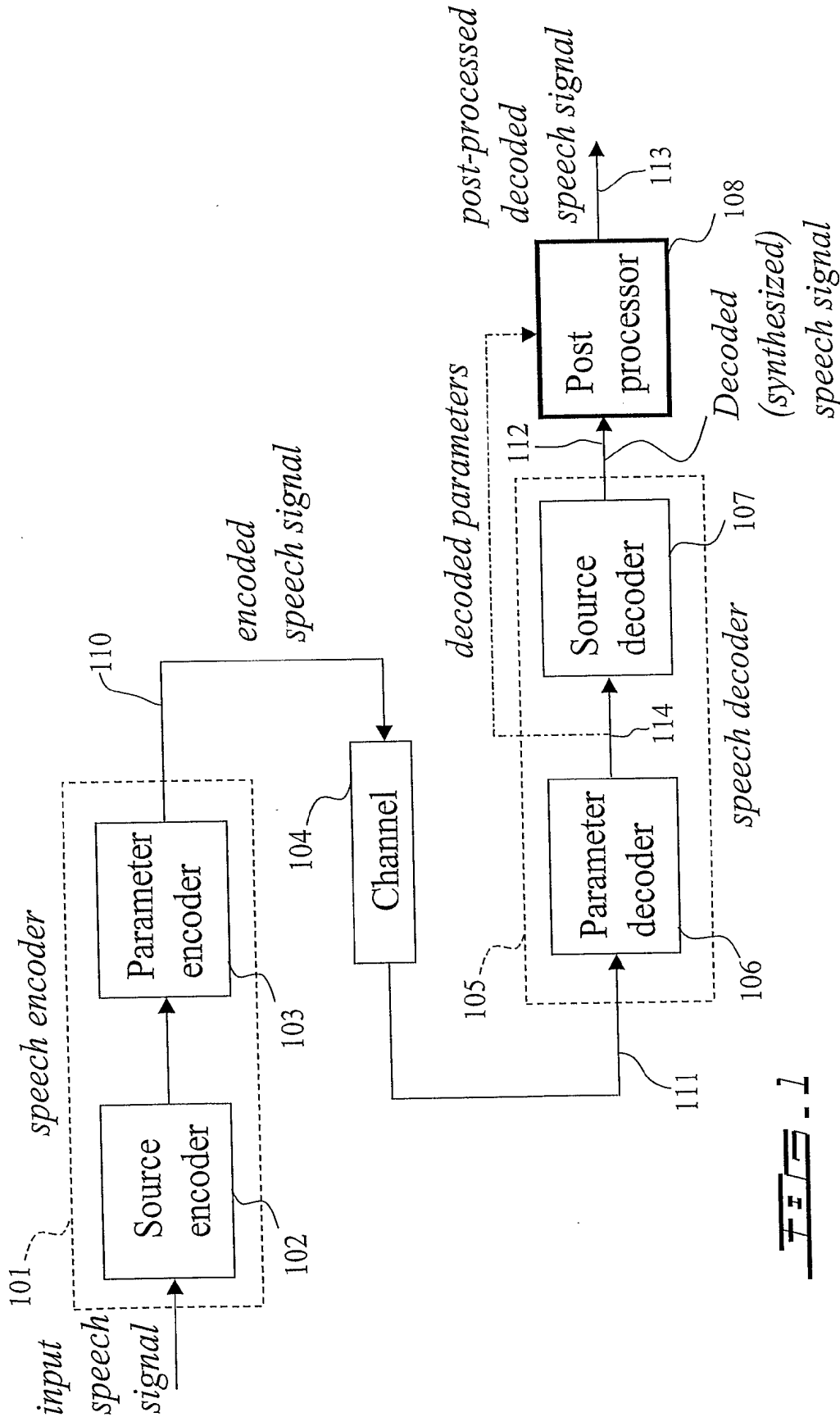
where  $x[n]$  is the decoded sound signal,  $y[n]$  is the pitch enhanced decoded

sound signal in a given sub-band,  $T$  is a pitch delay of the decoded sound signal, and  $\alpha$  is a coefficient varying between 0 and 1 to control an amount of inter-harmonic attenuation of the decoded sound signal.

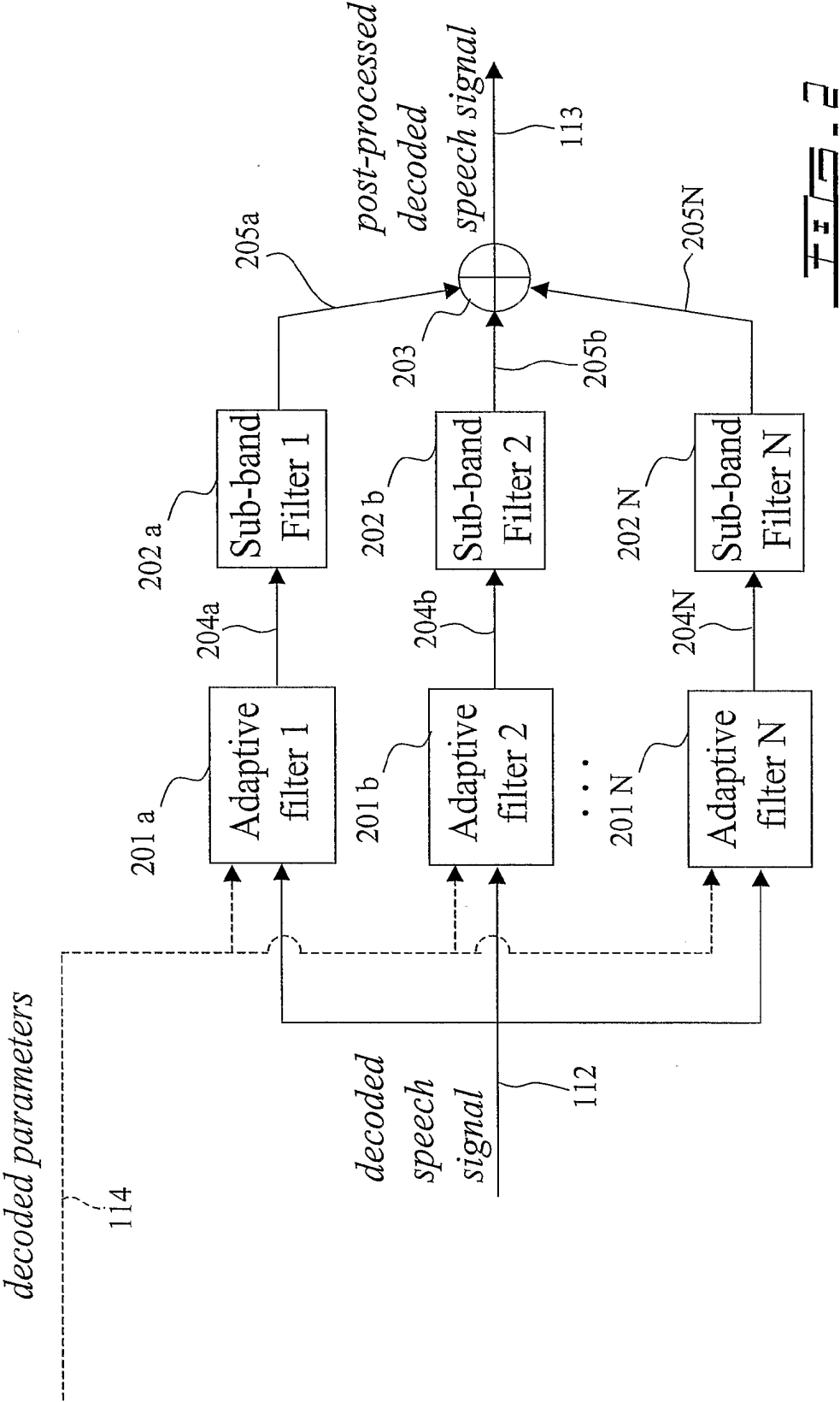
- 5           61. A post-processing device as defined in claim 32, wherein:  
the dividing means comprises means for dividing the decoded sound signal into a frequency upper-band signal and a frequency lower-band signal;  
and  
the post-processing means comprises means for post-processing the  
10 frequency lower-band signal.

62. A post-processing device as defined in claim 32, wherein the post-processing means comprises:  
means for determining a pitch value of the decoded sound signal;  
15 means for calculating, in relation to the determined pitch value, a high-pass filter with a cut-off frequency below a fundamental frequency of the decoded sound signal; and  
means for processing the decoded sound signal through the calculated high-pass filter.

- 20           63. A sound signal decoder comprising:  
an input for receiving an encoded sound signal;  
a parameter decoder supplied with the encoded sound signal for decoding sound signal encoding parameters;  
25 a sound signal decoder supplied with the decoded sound signal encoding parameters for producing a decoded sound signal; and  
a post processing device as recited in any of claims 32 to 62 for post-processing the decoded sound signal in view of enhancing a perceived quality of said decoded sound signal.







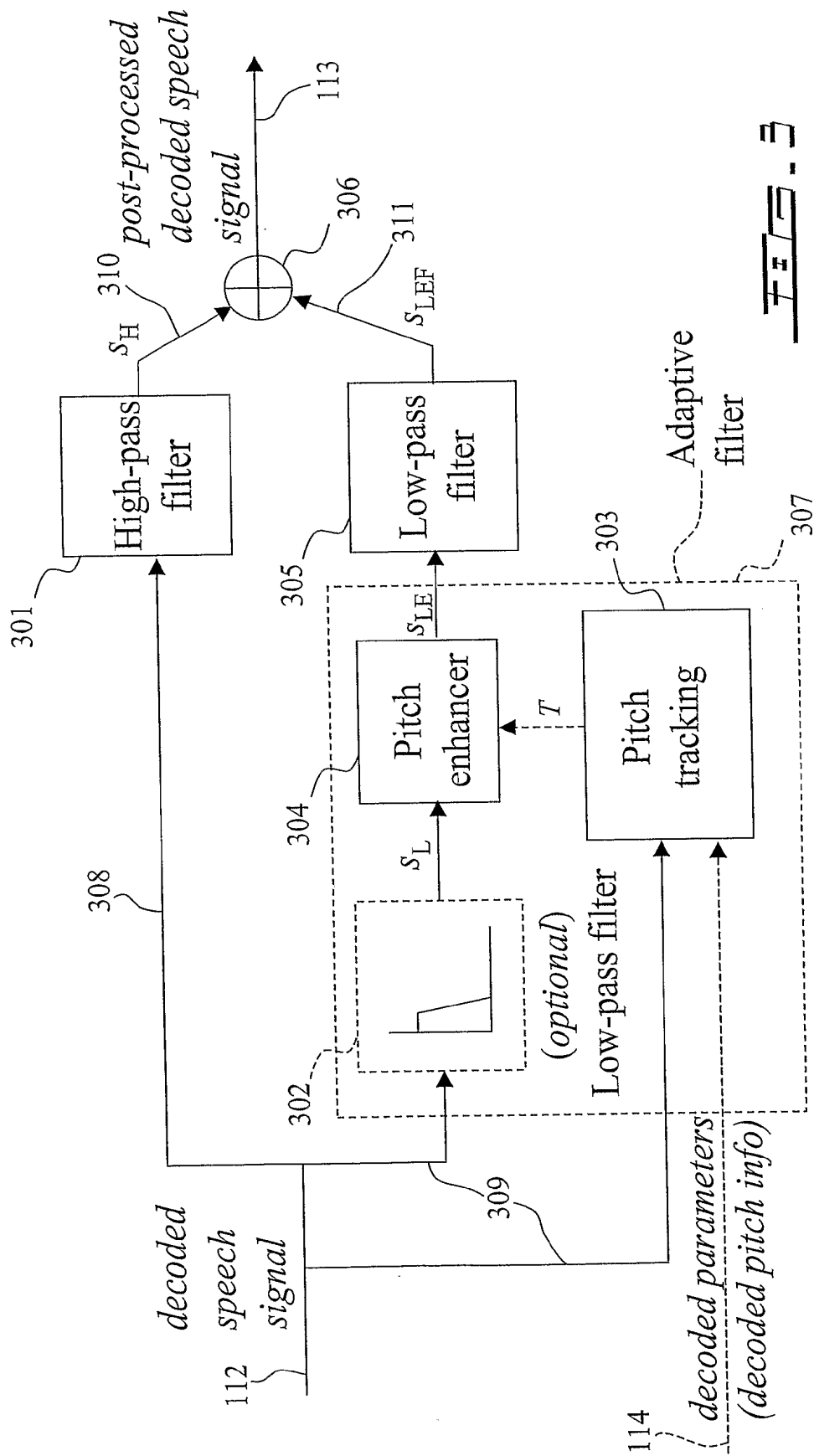


Fig. 3

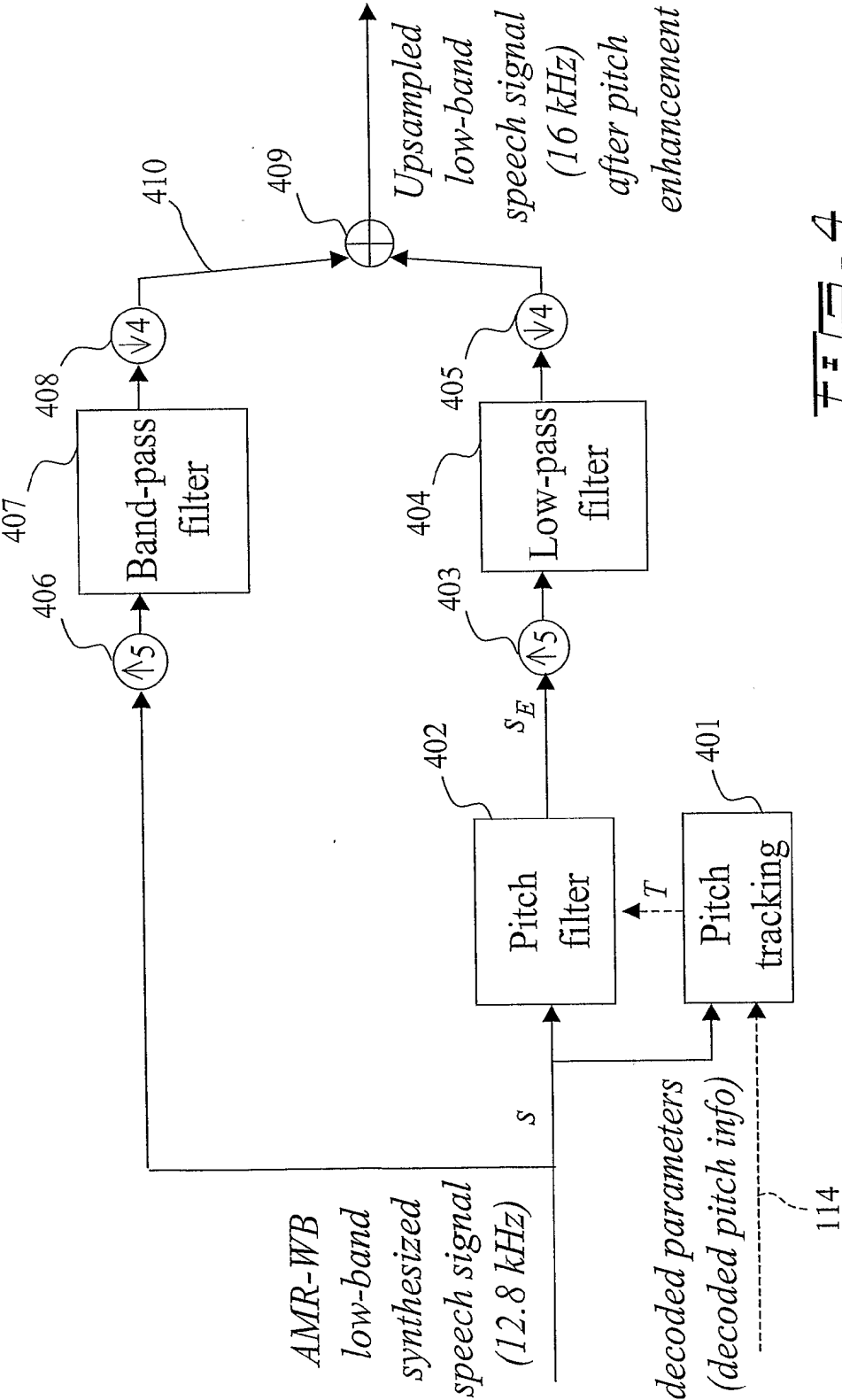


FIG. 4

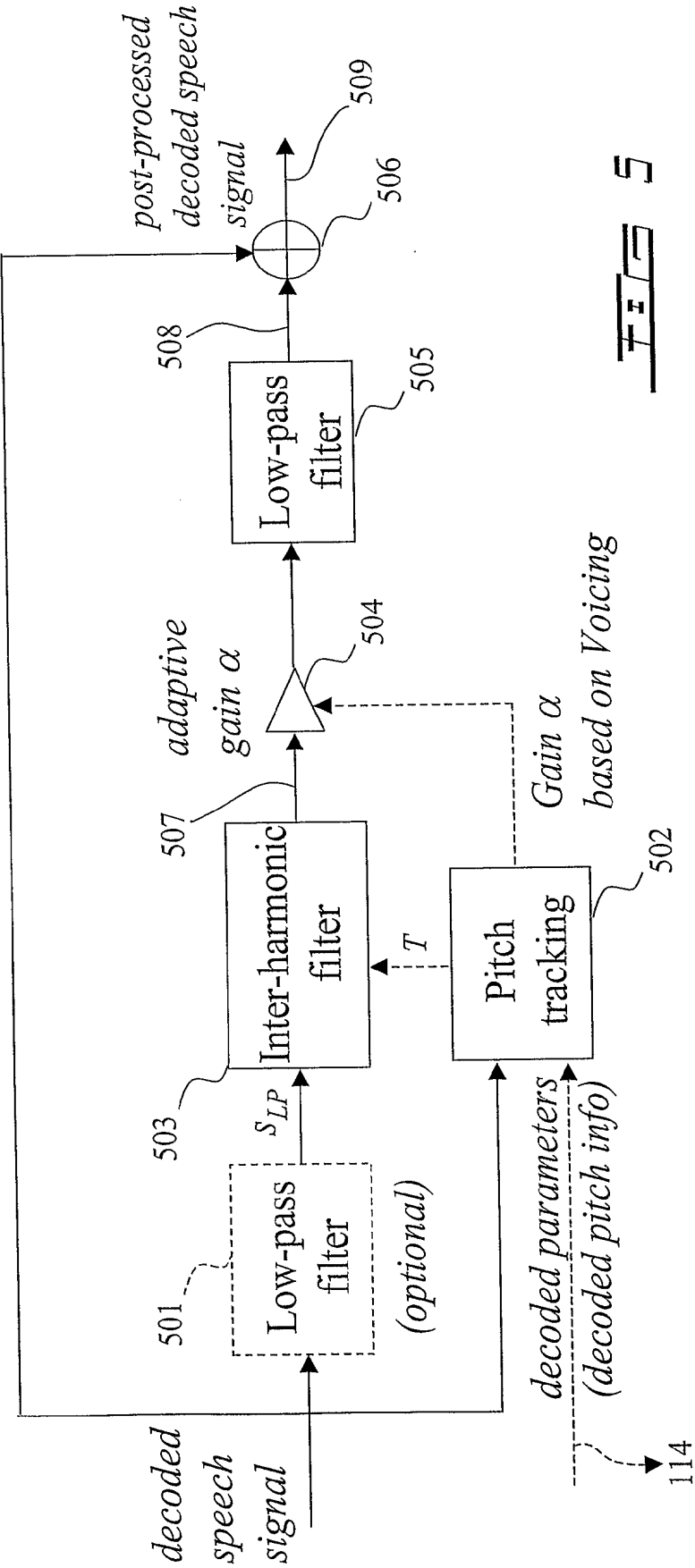
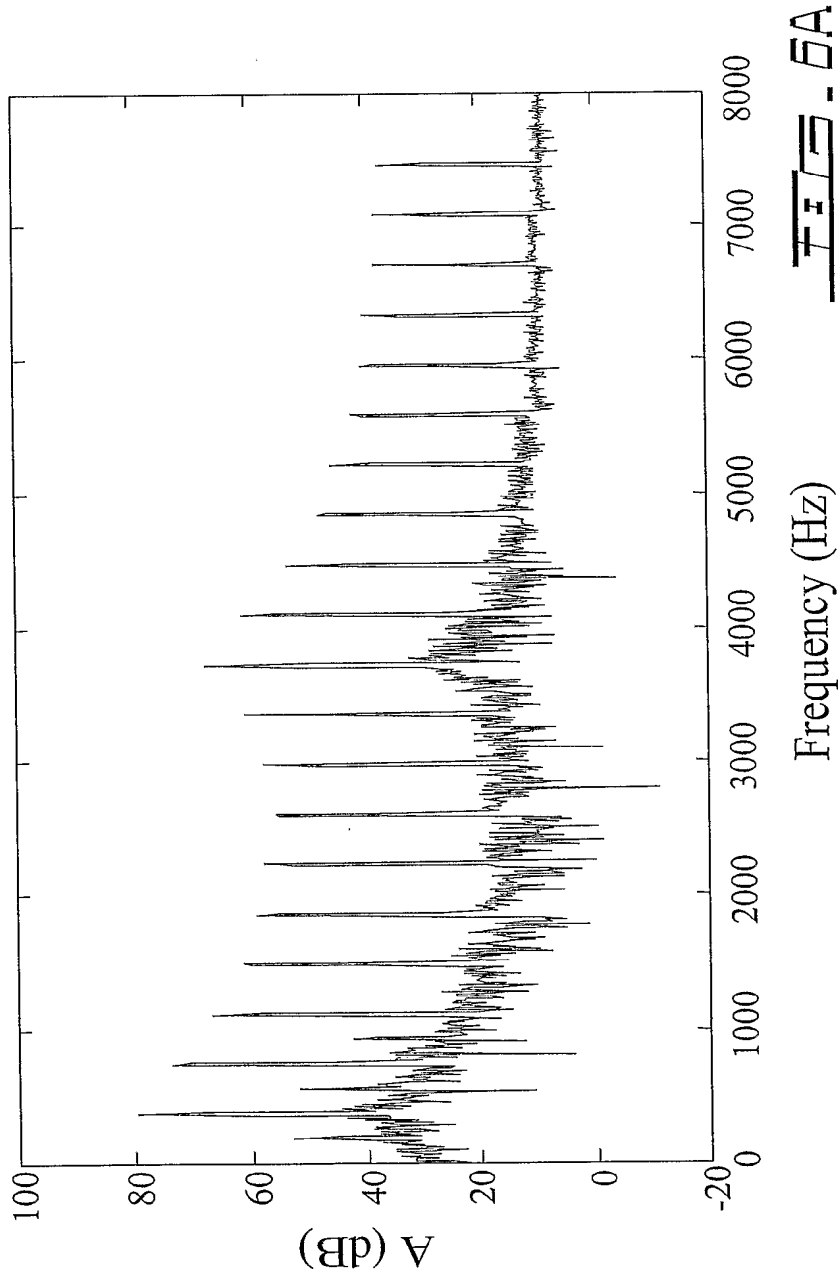
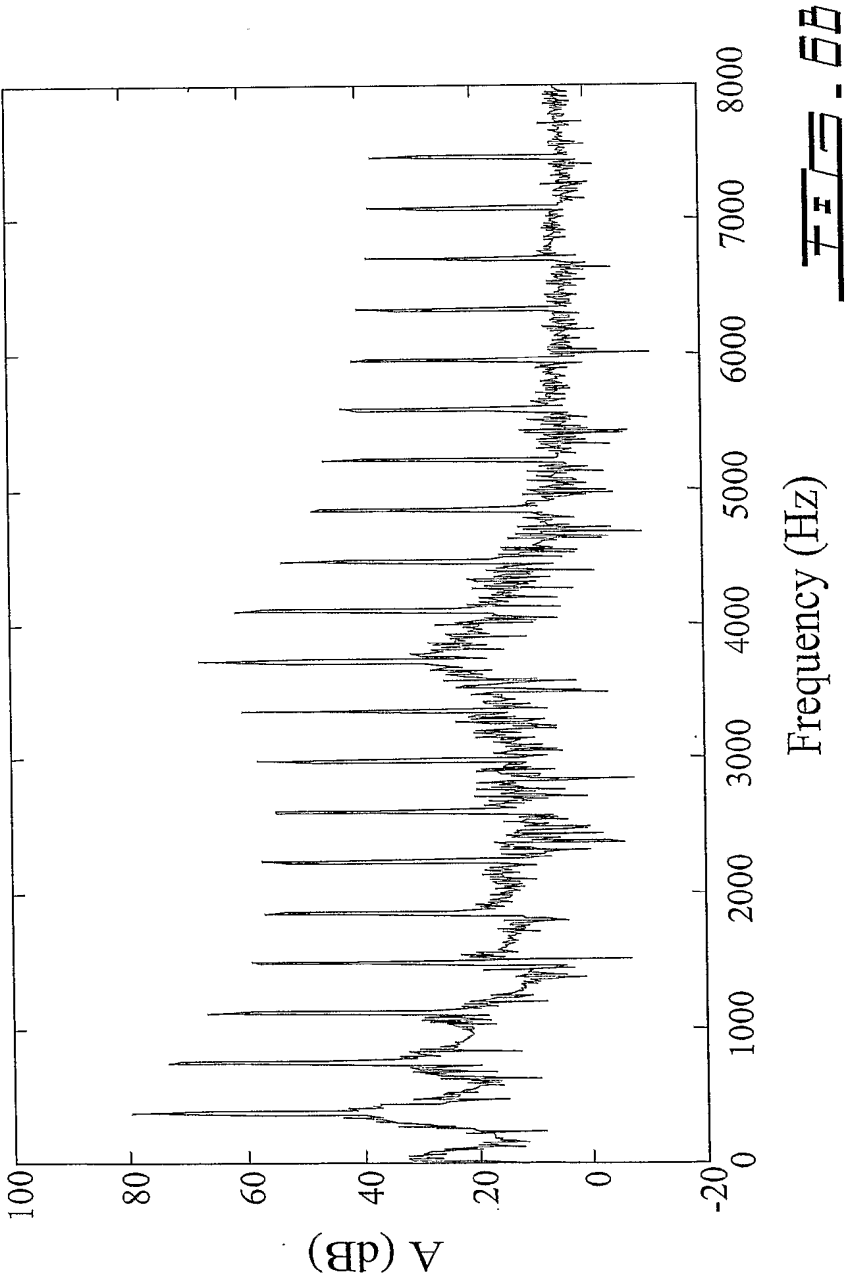


FIG 5





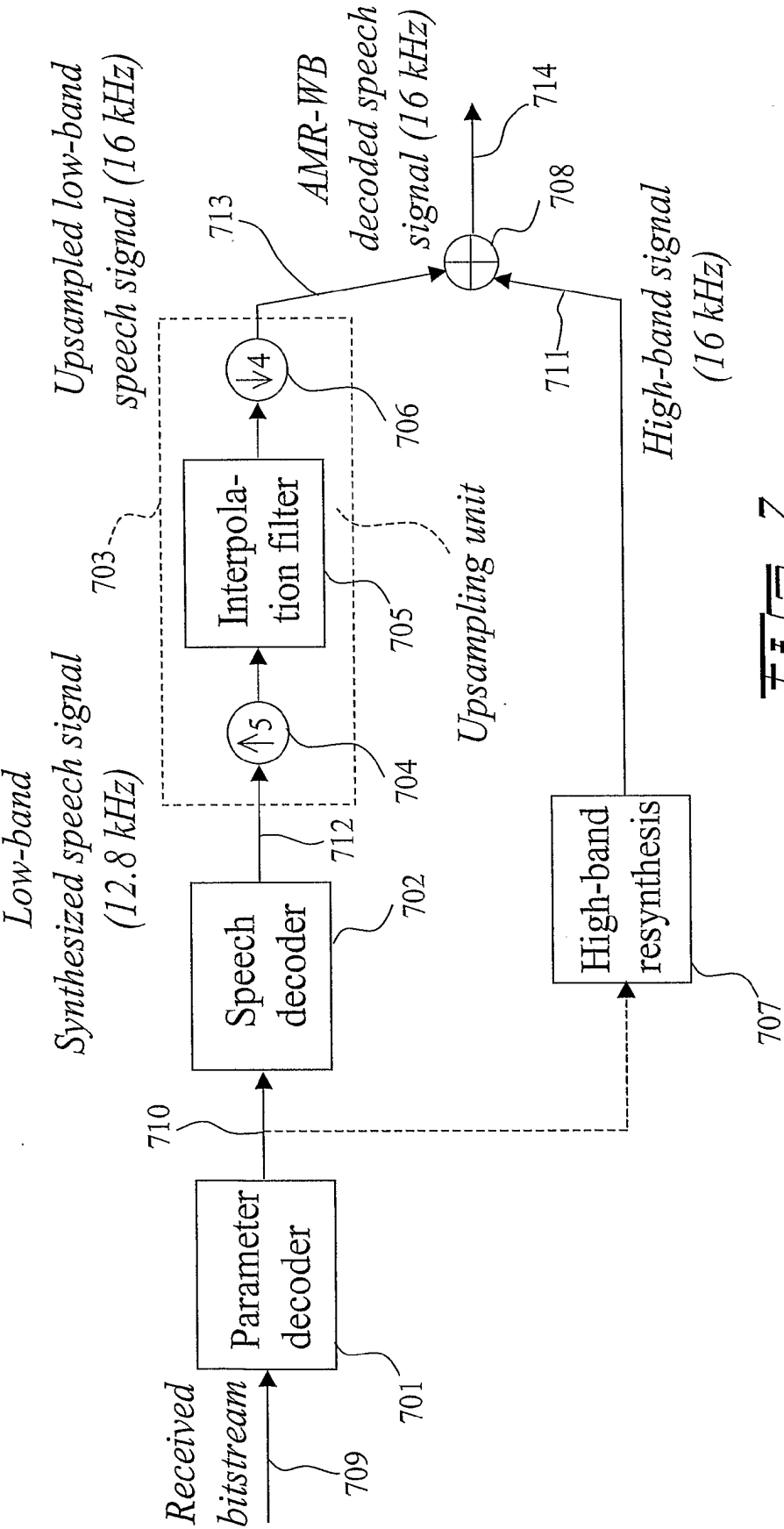
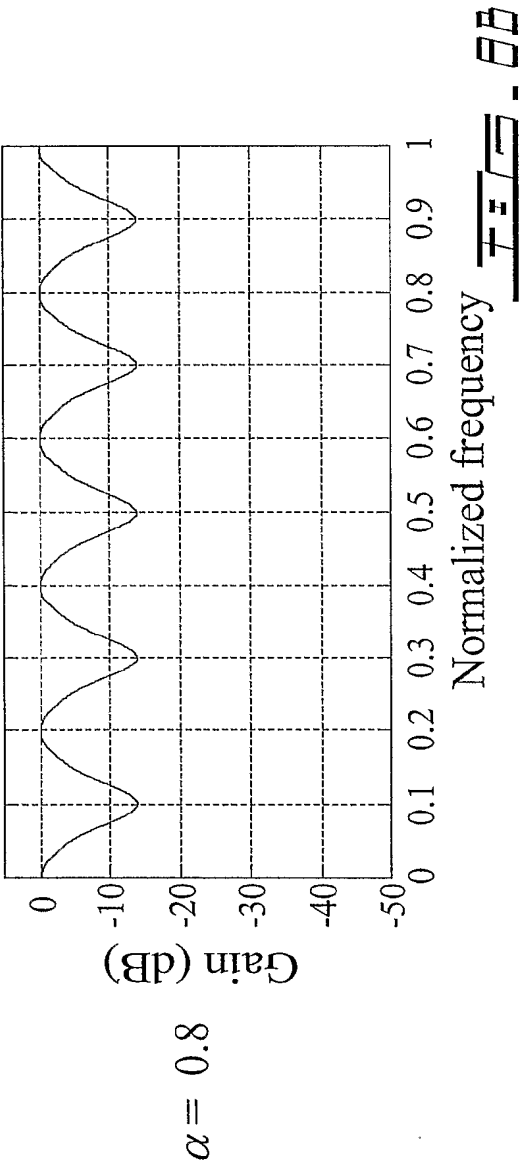
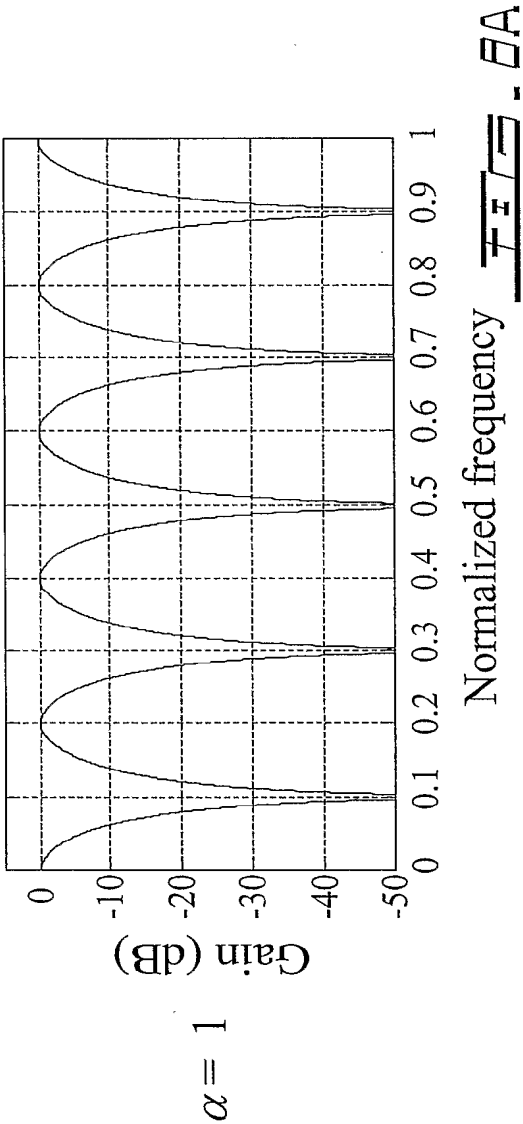
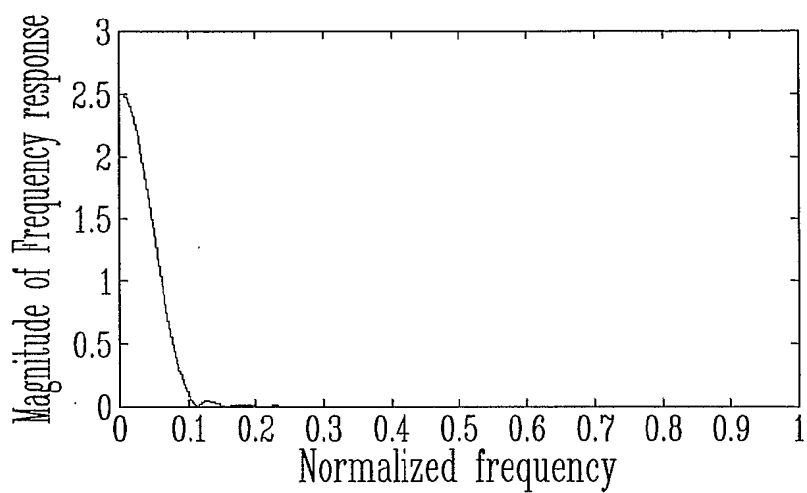
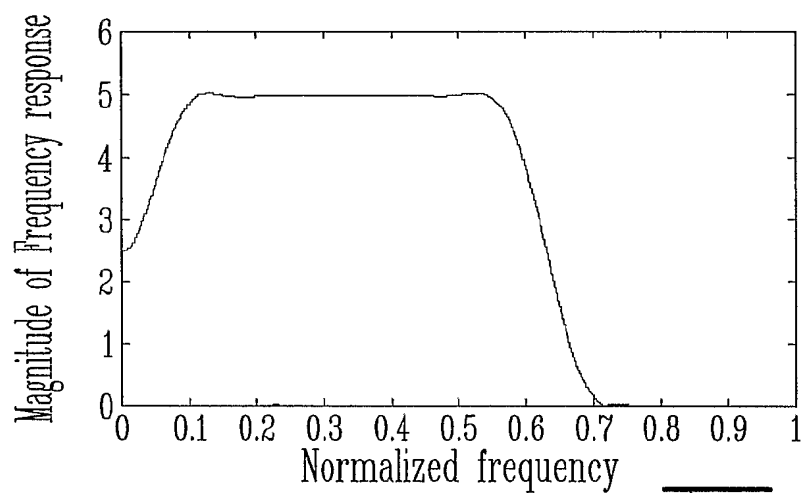
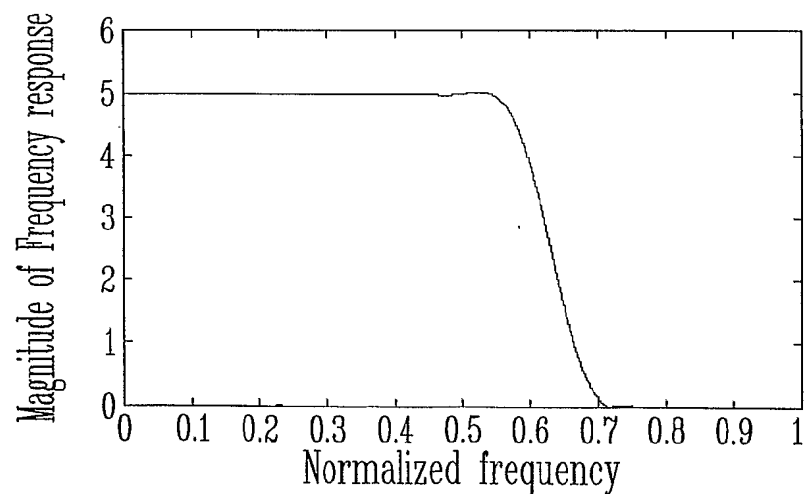


FIG. 7





FIG. 9AFIG. 9BFIG. 9C

